Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung
Certificate
Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. **Patent application No.** **Demande de brevet n°**

97120879.8

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Europäisches Patentamt | European Patent Office | Office européen des brevets

Anmeldung Nr:
Application no.: 97120879.8
Demande no:

Anmeldetag:
Date of filing: 27.11.97
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

MAX-PLANCK-GESELLSCHAFT ZUR FÖRDERUNG DER WISSENSCHAFTEN E.V.

14195 Berlin
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Improved method for the identification and characterization of interacting molecules by automated interaction mating

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12Q1/68

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

Max-Planck-Gesellschaft zur Förderung
der Wissenschaften e.V.
Our Ref.: B 3133 EP

IMPROVED METHOD FOR THE IDENTIFICATION AND CHARACTERIZATION OF INTERACTING MOLECULES BY AUTOMATED INTERACTION MATING

The present invention relates to an improved method for the identification and optionally the characterization of interacting molecules designed to eliminate the rather large numbers of false positive clones isolated by conventional two-hybrid systems. The method of the invention relies on a novel combination of selection steps used to detect clones that express interacting molecules. The present invention further relates to a kit useful for carrying out the method of the invention. The present invention provides for parallel, high-throughput or automated interaction screens for the reliable identification of interacting molecules by interaction mating.

Protein-protein interactions are essential for nearly all biological processes like replication, transcription, secretion, signal transduction and metabolism. Classical methods for identifying such interactions like co-immunoprecipitation or cross-linking are not available for all proteins or may not be sufficiently sensitive. Said methods further have the disadvantage that only by a great deal of energy, potentially interacting partners and corresponding nucleic acid fragments or sequences may be identified. Usually, this is effected by protein sequencing or production of antibodies, followed by the screening of an expression-library.

An important development for the convenient identification of protein-protein interactions was the yeast two-hybrid (2H) system presented by Fields and Song (1989). This genetic procedure not only allows the rapid demonstration of in vivo interactions, but also the simple isolation of corresponding nucleic acid sequences encoding for the interacting partners. The yeast two-hybrid system makes use of the features of a wide variety of eukaryotic transcription factors which carry two separable functional domains: one DNA binding

domain as well as a second domain which activates the RNA-polymerase complex (activation domain). In the classical 2H system a so-called "bait" protein comprising of a DNA binding domain (GAL4bd or lex A) and a protein of interest „X" are expressed as a fusion protein in yeast. The same yeast cell also simultaneously expresses a so called "fish" protein comprising an activation domain (GAL4ad or VP16) and a protein „Y". Upon the interaction of a bait protein with a fish protein, the DNA binding and activation domains of the fusion proteins are brought into close proximity and the resulting protein complex triggers the expression of the reporter genes, for example, HIS3 or lacZ. Said expression can be easily monitored by cultivation of the yeast cells on selective medium without histidine as well as upon the activation of the lacZ gene. The genetic sequence encoding, for example, an unknown fish protein, may easily be identified by isolating the corresponding plasmid and subsequent sequence analysis. Meanwhile, a number of variants of the 2H system have been developed. The most important of those are the "one hybrid" system for the identification of promoter binding proteins and the "tri-hybrid" system for the identification of RNA-protein-interactions (Li and Hershowitz, 1993; SenGupta et al., 1996; Putz et al., 1996).

The classical 2H system for the identification of protein-protein-interaction, has, until today, only been carried out on a laboratory scale. The various steps of this system need to be conducted serially. They are, therefore, quite time consuming. As a consequence, the 2H system has so far proven unsuitable for the large scale analysis of eukaryotic library vs library screens to investigate protein-protein networks. Although recent developments have taken into account these disadvantages (Bartel et al.,1996), a successful large scale search of interacting proteins, for example on the basis of a eukaryotic library vs. library screen, has not been reported. More importantly, also all of the so far developed 2H systems suffer from the serious drawback that many false-positive clones not representing any interactions between binding partners are isolated. This is particularly inconvenient in cases where large numbers of clones are to be analyzed because in the case of a eukaryotic library vs library screen typically several hundreds of thousands of clones have to be analyzed for the investigation of protein-protein networks.

The technical problem underlying the present invention was therefore to overcome these prior art difficulties and to furnish a system that reliably produces clones that express interacting molecules. This system should, moreover, be suitable for large-scale library vs library screens using a parallel, high-throughput or automated approach.

The solution to said technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to a method for the identification of at least one member of a pair or complex of interacting molecules, comprising:

(a) providing at least two sets of host cells each containing at least one genetic element with a selectable marker different for each set of host cells, said genetic elements each comprising genetic information specifying one of said molecules, said host cells further carrying a readout system that is activated upon the presence of autoactivating molecules;

(b) selecting against molecules that auto-activate said readout system by transferring progeny of at least one set of host cells to:

(ba) a selective medium which allows growth of said host cells in the presence of said selectable marker different for each set of host cells and which precludes growth of said host cells upon auto-activation of said readout system; or

(bb) a selective medium which allows growth of said host cells and visual differentiation between those cells whose readout system has been auto-activated and those whose readout system has not been auto-activated;

(c) combining in host cells said genetic elements from at least two different sets of host cells, wherein at least one set of host cells grows on said selective medium specified in (ba) or does not auto-activate said readout system on said selective medium specified in (bb);

(d) allowing at least one interaction, if any, to occur;

(e) identifying host cells obtained in step (c) or (d) containing interacting molecules that activate said readout system:

(f) identifying at least one member of said pair or complex of interacting molecules.

Preferably, said interaction is a specific interaction.

The terms "identification" and "identifying", as used in accordance with the present invention, relate to the ability of the person skilled in the art to detect positive clones that express interacting molecules due to the activation of the readout system in step (e) and optionally additionally to characterize at least one of said interacting molecules by one or a set of unambiguous features. Preferably, said molecules are characterized by the DNA sequence encoding them, upon nucleic acid hybridization or isolation and sequencing of the respective DNA molecules. Alternatively and less preferred, said molecules may be characterized by different features such as molecular weight, isoelectric point and, in the case of proteins, the N-terminal amino acid sequence etc. Methods for determining such parameters are well known in the art.

Preferably, said members specified by said genetic elements are connected to a further entity that will, upon the interaction, activate or contribute to the activation of said read out system. It is further preferred that said entity is conserved for each type of genetic element and that different types of genetic elements comprise different entities. It is additionally preferred that said member of said pair or complex of interacting molecules forms, when transcribed as RNA from said genetic element, an RNA transcript fused with RNA specifying said entity. Most preferably, said fused RNA transcript is translated to form a fusion protein comprising said member fused to said entity. As will be elaborated further herein below, said entity may be in one type of genetic element a DNA sequence encoding a DNA-binding domain and in a different type of genetic element a transactivating protein domain. Preferably, said genetic elements are vectors such as plasmids. The at least

two genetic elements comprised in said host cell are preferentially vectors from a library such as a cDNA or genomic library. Thus, the method of the invention allows the screening of a variety of host cells wherein the vector portion of said genetic elements is preferably the same for each type of genetic element whereas the potentially interacting molecules are representatives of a library and thus, as a rule and in case that the library has not been amplified, may differ in each host cell. In this connection the term "type of genetic element" refers to an element characterized by comprising the same entity, selectable and counterselectable markers.

Preferably, the "interaction" of said molecules is specific and characterized by a high binding constant. However, the term "interaction" may also refer to a binding between molecules with a lower binding constant which, however, must be sufficient to activate the readout system. The interaction that is detectable by the method of the invention preferably leads to the formation of a functional entity having a biological, physical or chemical activity which was not present in said host cell before said interaction occurred.

Said interaction may preferably lead to the formation of a functional transcriptional activator comprising a DNA-binding and a transactivating protein domain and which is capable of activating a responsive moiety that drives the activation of said readout system. For example, said moiety may be a promoter. Alternatively, said interaction may lead to a detectable fluorescence resonance energy transfer obtained by the interaction of fusion proteins containing, for example, the GFP type a and GFP type b fluorescent proteins (Cubbitt et al., 1995).

In a further embodiment, said interaction may lead to a detectable modification of a substrate by an enzyme such as a color reaction obtained by the cleavage of a propeptide by an enzyme. In all these embodiments of the invention, it is understood that the interacting molecules are preferably directly fused to the molecules driving the readout system.

The terms "autoactivate" or "autoactivation" relate to the fact that certain molecules encoded by said genetic elements are able to activate the readout system without the need for any interacting molecule. For example, the single fusion protein LexA-HIP1 is capable of activating the HIS3 and lacZ readout system without any corresponding interacting activation domain fusion protein.

The term "growth on selective media" refers to the fact that yeast cells containing one genetic element are placed on selective media that precludes growth of said cells upon autoactivation of said readout system, or the visual differentiation between cells whose readout system has been autoactivated and those cells whose readout system has not been autoactivated. For example, when a *ura3* yeast strain which contains a *URA3* reporter system and which also contains a plasmid expressing a LexA fusion protein that activates the *URA3* reporter system is selected on selective medium containing 5-fluoroorotic acid (5-FOA), the yeast cells can not grow on this medium because the *URA3* reporter system synthesises the enzyme orotidine-5'-phosphate decarboxylase that converts 5-FOA into the toxic compound 5-fluorouracil (Boeke et al., 1984). In contrast, on a selective medium lacking for example, tryptophan and which contains X-Gal, yeast cells that contain plasmids for the expression of LexA fusion proteins that either activate or do not activate the readout system can grow. However, the yeast cells in which the lacZ reporter system is activated will turn blue because the substrate X-Gal is cleaved into the coloured compound 5-bromo-4-chloro-indigo.

The term "growth on selective medium" also refers to the fact that host cells containing two genetic elements expressing interacting molecules which do not activate the readout system on their own, are selected on selective medium. For example, clones that express interacting LexA and GAL4ad fusion proteins which activate a URA3 and HIS3 reporter system can be selected on selective media lacking histidine and uracil. On this selective medium, only those yeast cells that contain interacting LexA and GAL4ad fusion proteins that activate the URA3 and HIS3 readout system can grow.

In accordance with the present invention it is envisaged that a counter selection against clones that express a single molecule able to activate the URA3 readout system can be carried out on culture media comprising 5-fluoroorotic acid. By applying this selection step prior to the preferably automated interaction mating, those clones that express autoactivating fusion proteins can be eliminated from a library of clones (Figure 1).

The URA3 readout system used for the elimination of false positive clones prior to the interaction mating is either present on an additional genetic element within the yeast cells or can be integrated into the yeast genome in a manner similar to the lacZ or HIS3 reporter genes. Figure 4 shows a plasmid that autonomously replicates in yeast and that contains the URA3 reporter construct which can be used for the elimination of autoactivating moleules.

Whereas it is only necessary that one set of host cells grows on said selective medium specified in step (ba) or does not auto-activate said readout system on said selective medium specified in step (bb), it is also advantageous to combine in said host cells genetic elements from at least two sets of host cells that fulfil these criteria.

The method of the present invention provides a highly effective tool for the elimination of false positive clones that have proven to dramatically reduce the overall usefulness of the two-hybrid system. The advantages associated with the method of the invention have a significant impact in particular on the number of clones that express potentially interacting partners that can conveniently be analysed. For example, even work on the laboratory scale will be more effective since positive clones that express interacting partners can be easily obtained because the clones expressing autoactivating fusion proteins are eliminated prior to the screeing of interacting molecules. In contrast, the elimination of autoactivating fusion proteins prior to the screening of interacting fusion proteins using the state of the art yeast two-hybrid system is labour intensive because replica plating and consideration of the growth patterns of clones on at least two different media has to be performed. Bartel et al. (1996) described a method for the elimination of false positives by

replica plating clones that express one fusion protein from SD-leu and SD-trp plates, to SD-his plates. Clones that showed growth on the SD-his plates where identified as false positives and were subsequently not used for interaction mating. With the method of the invention, by plating of the yeast cells directly on selective plates containing 5-FOA, positive clones that do not activate the readout system can be easily isolated without the need for replica plating. In the long run, it is expected that the method of the present invention is especially advantageous for automated analysis of a large number of yeast clones containing interacting molecules since many specific interactions and the individual members of these interactions can be identified in a parrallel and high thoughput approach.

A schematic overview of one embodiment of the method of the invention is provided in Figure 2 illustrating the invention: For the parallel analysis of a network of protein-protein interactions with the method of the invention, two libraries of plasmid constructs that express DNA binding domain or activation domain fusion proteins are provided. These libraries may consist of specific DNA fragments or a multitude of unknown DNA fragments ligated into binding domain and activating domain plasmids which contain the selectable markers TRP1 and LEU2, respectively. The libraries are transformed into either Mata or Matα yeast strains containing the novel URA3 reporter system and are subsequently plated onto selective media containing 5-fluoroorotic acid (5-FOA). Only those yeast cells that express fusion proteins unable to auto-activate the URA3 reporter system will grow in the presence of 5-FOA. The resulting yeast strains that express only non-auto-activating molecules can then be directly used in an automated interaction mating approach to generate ordered arrays of diploid strains which can be assayed for activation of the lacZ readout system.

a) Individual yeast cells that express single fusion proteins unable to activate the URA3 reporter system are transferred into wells of a 384-well microtitre plate using a picking robot. The yeast strains held in the microtitre plates can optionally be replicated and stored. The microtitre plates contain a growth medium lacking amino acids appropriate to maintain the corresponding plasmids in the yeast strains. The interaction matings are subsequently

performed by automatically transfering a *Mata* and a *Mat*α yeast strain to the same position on a Nylon membrane using automated systems as described by Lehrach *et al.* (1997). Alternatively, a pipetting or micropipetting system (Schober *et al.* 1993) can be used to transfer small volumes of individual liquid cultures of a yeast strain onto a planar carrier onto which a lawn of yeast cells derived from at least one yeast clone of the opposite mating type is sprayed or applied. By both methods ordered arrays of yeast clones are incubated overnight at 30°C to allow interaction mating to occur. The resulting diploid cells are then analysed in a ß-Gal assay as described by Breeden & Nasmyth (1985).

b) Yeast strains that grew on selective media containing 5-FOA are pooled and interaction mating between the *Mata* and *Mat*α strains is made within liquid YPD medium. Those diploid yeast strains that express interacting proteins are selected by plating on selective medium lacking histidine and uracil. The selective markers TRP1 and LEU2 maintain the plasmids in yeast strains grown on selective media. HIS3, URA3 and lacZ represent reporter genes in the yeast cells, which are expressed on activation by interacting fusion proteins. The readout system is, in the present case, growth on medium lacking histidine and/or uracil and enzymatic activity of ß-galactosidase which can be screened at a later time point. A colony picking robot is used to pick the diploid yeast colonies into individual wells of 384-well microtiter plates containing selective medium lacking histidine, and the resulting plates are incubated at 30°C to allow cell growth. The interaction library optionally may be replicated and stored. Using a spotting robot, diploid cells are transferred to replica membranes which are subsequently placed onto growth medium. The resulting regular arrays of diploid yeast clones are analysed for ß-Gal activity as described by Breeden & Nasmyth (1985). In either case a) and b), a digital image from each dried membrane is obtained with a CCD camera which is then stored on computer. Using digital image processing and analysis (Lehrach et al. 1997) clones that express interacting fusion proteins can be identified by considering the ß-Gal activity of these clones spotted in a defined pattern on the membrane. The individual members comprising the interactions can then be identified by one or more techniques,

including PCR, sequencing, hybridisation, oligofingerprinting or antibody reactions.

The genetic elements specified here and above may further and advantageously be equipped with at least two different selection markers functional in bacteria such as E.coli. Such selection markers, for example aphA (Pansegrau et al., 1987) or bla allow the easy separation of said genetic elements upon retransformation into E.coli strains.

In a preferred embodiment of the method of the present invention said pair or complex of interacting molecules is selected from the group consisting of RNA-RNA, RNA-DNA, RNA-protein, DNA-DNA, DNA-protein, protein-protein, protein-peptide, or peptide-peptide interactions.

Accordingly, the method of the invention is applicable in a wide range of biological interactions. For example, the invention will be useful in identifying peptide-protein or peptide-peptide-interactions by employing synthetic peptide libraries (Yang et al., 1995).
Two applications of interests are the application of a large scale two-hybrid system for the detection of protein-protein interactions involved in medically relevant pathways which may be useful as diagnostic or therapeutic targets for the treatment of disease, and a large scale tri-hybrid which is one example of said complex of interacting molecules mentioned herein above of, for example, a system for the identification of novel post-transciptional regulators and their binding sites (Putz et al., 1996; SenGupta et al., 1996). In this regard, it should be noted that a complex, in accordance with the invention, may comprise more than three interacting molecules. Such a complex may be composed of biologically and chemically different members.

For example, to identify interacting RNA binding proteins and RNA molecules, a plasmid expressing a LexA-HIV -1Rev protein, a plasmid comprising a RNA sequence in fusion with the Rev responsive element and a plasmid expressing a potentially RNA-interacting protein in fusion with an activation domain may be present in one cell. The plasmids encoding the RNA fusion

molecule and the activation domain fusion protein must contain different selectable and counterselectable markers. If the RNA fusion molecule interacts with the respective two fusion proteins, the readout system is activated. To test whether the RNA fusion molecule or the activation domain fusion protein interact, the method of the invention is used to investigate the activation of the readout system in the absence of these fusion molecules.

In a further preferred embodiment, said genetic elements are plasmids, artificial chromosomes, viruses or other extrachromosomal elements.

Whereas it is preferred, due to the easy handling, to employ plasmids that specify the genetic elements in accordance with the present invention, the persons skilled in the art will be able to devise other systems that carry said genetic elements and that are identified above.

In an additional preferred embodiment, said readout system is a detectable protein. A number of readout systems are known in the art and may, if necessary, be adapted to be useful in the method of the invention.

Most preferably, said detectable protein is that encoded by the gene lacZ, HIS3, URA3, CAN1, LYS2, CYH2, sacB or HPRT, respectively. As is well known in the art, the expression of the ß-gal enzyme in yeast can be used for the formation of a detectable blue colony after incubation in X-Gal solution. Of course, the method of the invention is not restricted for use of only one readout system. On the contrary, if desired, a number of such readout systems may be combined. Said combination of the number of readout systems is, in accordance with the present invention, also comprised by the term "readout system". Such a combination will provide an additional safe guard for the identification of clones containing interacting partners.

Although the two-hybrid system has been developed in yeast, the method of the invention can be carried out in a variety of host systems. Preferred of those are yeast cells, bacterial cells, mammalian cells (Wu et al. 1996), insect cells or plant cells. Preferably, the bacterial cells are E. coli cells.

Of course, the genetic elements may be engineered and prepared in one host organism and then, e.g. by employing shuttle vectors, be transferred to a different host organism where it is employed in the method of the invention.

In another preferred embodiment, the method of the present invention comprises transforming or transfecting said sets of host cell with at least one of said genetic elements prior to step (a).

Whereas the person skilled in the art may initiate the identification method of the invention starting from fully transformed or transfected sets of host cells, he may wish to first generate such host cells in accordance with the aim of his research or commercial interest. For example, he may wish to generate a certain type of library first that he intends to screen against a second library already present in said host cells. Alternatively, he may have in mind to generate two different libraries that he wants to screen against each other. In this case, he would need to first transform said sets of host cells with either type of genetic element.

In another preferred embodiment, said host cells with said genetic elements are generated by cell fusion, conjugation or interaction mating.

In an additional preferred embodiment of the method of the invention, said visual differentiation in step (bb) is based on a color reaction which is induced by said readout system. For example, the readout system may comprise an enzyme that cleaves a substrate giving rise to a colour reaction.

It is further preferred in accordance with the present invention that said selectable markers are auxotrophic or antibiotic markers.

It is important to note that some of the markers that are used as a readout system, may also be used as selectable markers. It is further important to note that one and the same marker can not be used as selectable marker and as part of the readout system at the same time.

Most preferably, said auxotrophic or antibiotic markers are selected from LEU2, TRP1, URA3, HIS3, ADE2, LYS2 and Zeocin.

In a further preferred embodiment of the method of the invention, said selective medium specified in step (ba) comprises 5-fluoroorotic acid, canavanine, cycloheximide or α-amino adipate.

Planning of experiments may require that the process need not be done continuously. In such cases, the researcher may wish to store the host cells for further use. Accordingly, a further preferred embodiment of the invention relates to a method wherein progeny of host cells obtained in step (b) or (d) are transferred to a storage compartment.

In particular in cases where a large number of clones is to be analyzed, said transfer is advantageously effected or assisted by automation or a picking robot. How such a picking robot may actually be put into practice, is described for example in Lehrach et al. (1997). Naturally, other automation or robot systems that reliably pick progeny of said host cells into predetermined arrays in the storage compartments may also be employed. However, if selection against false positive clones is to be made by visual differentiation according to step (bb), it is advantageous to use an automated system that can distinguish between the visual characteristics of clones expressing auto-activating and non auto-activating molecules.

The host cells will, in this embodiment, be propagated in said storage compartment and provide further progeny for the additional tests. Preferably, replicas of said storage compartment maintaining the array of clones are set up. Said storage compartments comprising the host cells and the appropriate media may be maintained in accordance with conventional cultivation protocols. Alternatively, said storage compartments may comprise an anti-freeze agent and therefore be appropriate for storage in a deep-freezer. This embodiment is particularly useful when the evaluation of potential interacting partners is to be postponed. As is well known in the art, frozen host cells may

easily be recovered upon thawing and further tested in accordance with the invention.

Most preferably, said anti-freeze agent is glycerol which is preferably present in said media in an amount of 3 - 25% (vol/vol).

In a further particularly preferred embodiment of the method of the invention, said storage compartment is a microtiter plate. Most preferably, said microtiter plate comprises 384 wells. Microtiter plates have the particular advantage of providing a pre-fixed array that allows the easy replicating of clones and furthermore the unambiguous identification and assignment of clones throughout the various steps of the experiment. The 384 well microtiter plate is, due to its comparatively small size and large number of compartments, particularly suitable for experiments where large numbers of clones need to be screened.

Depending on the design of the experiment, the host cells may be grown in the storage compartment such as the above microtiter plate to logarithmic or stationary phase. Growth conditions may be established by the person skilled in the art according to conventional procedures. Cell growth is usually performed between 15 and 45 degrees Celsius.

Combining in host cells said genetic elements from at least two different sets of host cells in step (c) is advantageously conducted by transfering host cells to a planar carrier which is subsequently placed on growth media to allow interaction mating to occur. Alternatively, said transfer is made to a planar carrier already placed on the growth medium or said transfer may be made directly to the growth medium.

Most preferably, said transfer is made by automation or by using a spotting robot or by using a pipetting or micropipetting device, both of which deliver at least one clone of each mating type to the same position or in close proximity to each other allowing interaction mating. How such a spotting robot may be devised and equipped is, for example, described in Lehrach et al. (1997).

Naturally, other automation or robotic systems that reliably create ordered arrays of clones may also be employed.

Alternatively, an automated system can be used to transfer only host cells of a single mating type to a planar carrier and then host cells of the opposite mating type are applied onto the planar carier by another method to allow interaction mating. For example, a spotting robot as described by Lehrach et al (1997) or a microdispensing devise (Schober et al., 1993) can be used to transfer yeast cells of mating type-a to a planar carrier, onto which then a suspension of yeast clones of the opposite mating type is sprayed. Interaction mating is then conducted by placing said planar carrier onto growth medium and subsequent transfer to selective medium.

Most advantageously, said transfer is effected in a regular grid pattern at densities of 1 to 1000 clones per square centimeter. The progeny of said host cells may be transferred to a variety of planar carriers. Most preferred is a membrane which may, for example, be manufactured from Nylon, nitro-cellulose or PVDF.

The growth medium to allow interaction mating may be in liquid or in solid form. Preferably, said medium when used in conjunction with a spotting robot and membranes as planar carriers is solidified agar on which said spotted membranes are subsequently placed. Alternatively, and also preferably, said medium when liquid is held within microtitre plates and said transfer is made by replication. However, it should be recognised that a person skilled in the art will be able to conduct interaction mating within liquid medium using pooled sets of host cells. In this case, and as indicated in Figure 2 route b, host cells that express interacting molecules can be selected using an appropriate selective medium. For example, to select for yeast cells that express interacting fusion proteins that activate a HIS3 reporter system, clones are plated on selective medium lacking histidine, and only clones that express interacting fusion proteins will grow. In this embodiment of the invention, it is most preferable to transfer said host cells that grow on said selective media to a storage compartment, most preferably with a picking robot such as described by Lehrach et al. (1997). It is a further preferred

embodiment that said diploid host cells held in the storage compartment are transfered to a planar carrier using a spotting robot as described by Lehrach et al (1997). The activation state of the readout system in said clones when grown on said planar carrier can be evaluated.

Referring now to the step (e) of the method of the invention, the readout system can be analyzed by a variety of means. For example, it can be analyzed by visual inspection, radioactive, chemiluminescent, fluorescent, photometric, spectrometric, infra red, colourimetric or resonant detection.

Preferably, said identification of host cells that express interacting fusion proteins is effected by visual means from consideration of the activation state of said readout system of clones containing said at least two genetic elements resulting from step (c) and after allowing at least one interaction, if any, to occur in step (d).

Also preferably, said identification of host cells that express interacting fusion proteins in step (e) is effected or assisted by digital image analysis or processing. In this embodiment, positive clones which are preferably arrayed on a planar carrier such as a membrane are identified by inspection of digital images obtained from the membrane after activation of said readout system from clones on said planar carrier.

Most preferably, the identity of positive host cells are stored on computer, for example within a relational database.

Identification of the at least one member of the pair or complex of interacting molecules may be effected by a variety of means. For example, molecules can be characterized by nucleic acid hybridization, oligonucleotide hybridization, nucleic acid or protein sequencing, restriction digestion, spectrometry or antibody reaction. Once the first member of an interaction has been identified, the second member or further members can also be identified by any of the above methods. Preferably the identification of at least one

member of an interaction is effected by nucleic acid hybridization, antibody binding or nucleic acid sequencing.

If nucleic acid hybridization is to be carried out, the nucleic acid molecules comprised in the host cell and encoding for at least one of the interacting molecules is preferably affixed to a planar carrier. As is well known in the art, said planar carrier to which said nucleic acid may be affixed, can be for example, a Nylon-, nitrocellusose- or PVDF membrane, glass or silica substrates (DeRisi et al. 1996; Lockhart et al. 1996). Said host cells containing said nucleic acid may be transferred to said planar carrier and subsequently lysed on the carrier and the nucleic acid released by said lysis is affixed to the same position by appropriate treatment. Alternatively, progeny of the host cells may be lysed in a storage compartment and the crude or purified nucleic acid obtained is then transferred and subsequently affixed to said planar carrier. Advantageously, said nucleic acids are amplified by PCR prior to transfer to the planar carrier. Most preferably said nucleic acid is affixed in a regular grid pattern in parallel with additional nucleic acids representing different genetic elements encoding interacting molecules. As is well known in the art, such regular grid patterns may be at densities of between 1 and 50,000 elements per square centimeter and can be made by a variety of methods. Preferably, said regular patterns are constructed using automation or a spotting robot such as described in Lehrach et al. (1997) and Maier et al. (1997) and furnished with defined spotting patterns, barcode reading and data recording abilities. Thus it is possible to correctly and unambiguously return to stored host cells containing said nucleic acid from a given spotted position on the planar carrier. Also preferably, said regular grid patterns may be made by pipetting systems, or by microarraying technologies as described by Shalon et al. (1996), Schober et al (1993) or Lockart et al. (1996). Identification is, again, advantageously effected by nucleic acid hybridisation.

Using a detectable nucleic acid probe of interest, homologous nucleic acids which are affixed on the planar carrier can be identified by hybridization. From the spotted position of said homologous identified nucleic acid on the planar carrier, the corresponding host cell in the storage compartment can be

identified which contains at least one member of the interaction. The second or any further member of the interaction can now be identified by any of the above methods. For example, by use of a radioactively labeled Ras probe, homologous nucleic acids on the planar carrier can be identified by hybridization. The Ras interacting proteins can now be identified from the corresponding host cell that contains both the first genetic element homologous to the Ras probe and the second genetic element encoding for these Ras interacting proteins.

If multiple oligonucleotide hybridizations are carried out on the nucleic acids affixed to the planar carrier, oligofingerprints of all genetic elements encoding the interacting proteins can be obtained. These oligofingerprints can be used to identify all members of the interactions or those members that belong to specific gene families, as described in Maier et al. (1997).

Advantageously, the nucleic acid molecules that encode the interacting proteins are, prior to identification such as by DNA sequencing, amplified by PCR or in said genetic elements in host cells and preferable in *E. coli*. Amplification of said genetic elements is conducted by multiplication of the *E. coli* cells and isolation of said genetic elements. Methods of identifying the nucleic acids that encode interacting proteins by DNA sequencing and analysis are well known in the art. By amplifying and sequencing the nucleic acids that encode all members of an interaction from the same clone, the identity of all members of the interaction can be determined.

If a specific antibody is to be used to determine whether a protein of interest is expressed as a fusion protein within an interaction library, it is advantageous to affix all fusion proteins expressed from the interaction library on to a planar carrier. For example, clones of the interaction library that express fusion proteins can be transferred to a planar carrier using a spotting robot as described in Lehrach et al (1997). The clones are subsequently lysed on the carrier and released proteins are affixed onto the same position. Using, for example, an anti-HIP1-antibody (Wanker et al. 1997), clones from the interaction library that contain HIP1 fusion proteins and an unknown

interacting fusion protein can be identified. The unknown member of the interacting pair of molecules can now be identified from the corresponding host cell by any of the above methods. The antibodies used as probes may be directly detectably labeled. Alternatively, said antibodies may be detected by a secondary probe or antibody which may be specific for the primary antibody. Various alternative embodiments using, for example, tertiary antibodies may be devised by the person skilled in the art on the basis of his common knowledge.

Most advantageously, when said identification of members comprising an interaction is effected using said regular grids, a digital image of the planar carrier after hybridization or antibody reaction is obtained and analysis is effected by digital image storage, processing or analysis using an automated or semi-automated image analysis system, such as described in Lehrach et al. (1997).

Most preferably, the information comprising the identity of the host cell and the identity of the interacting molecules expressed by the genetic elements contained within the host cell are stored on a computer, for example within a relational database.

These data are available for the establishment of a network of interactions. By collecting the information from a whole interaction library, the inter-relationship between many different interacting molecules can be determined and thus enable the establishment of a network of interactions. Preferably, said data can be accessed though the use of software tools or graphical interfaces that enable the investigator to easily query the established interaction network with a biological question or to develop the established network by the addition of further data.

Advantageously, those molecules identified as interacting with many different molecules can be recorded. This information can reduce the work needed to further characterize particular interactions since those interactions comprising

of a molecule found to interact with many other molecules within the yeast two-hybrid system may be suspected of being artifactual (Bartel et al., 1993).

A significant advantage of the method of invention over existing yeast two-hybrid systems is the scale at which such identification of interactions and interaction members can be made. Preferably, the method of invention screens library vs library interactions using arrayed interaction libraries. Thus, the method of invention allows, in an efficient manner, a more complete and exhaustive generation of protein-protein interaction networks than existing methods. An established and exhaustive network of protein-protein interactions is of use for many purposes as shown Figure 3. For example, it may be used to predict the existence of new biological interactions or pathways, or to determine links between biological networks. Furthermore with this method, the function and localization of previously unknown proteins can be predicted by determining their interaction partners. It also can be used to predict the response of a cell to changes in the expression of particular members of the networks. Finally, these data can be used to identify proteins or interactions between proteins within a medically relevant pathway which are suitable for therapeutic intervention, diagnosis or the treatment of a disease.

The present invention also relates to a method for the production of a pharmaceutical composition comprising formulation said at least one member of interacting molecules identified by the method of the invention in a pharmaceutically acceptable form.

Said pharmaceutical composition comprises at least one of the aforementioned compounds isolated by the method of the invention, either alone or in combination, and optionally a pharmaceutically acceptable carrier or exipient. Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions etc. Compositions comprising such carriers can be formulated by conventional methods. These pharmaceutical compositions can be administered to subject

in need thereof at a suitable dose. Administration of the suitable compositions may be effected by different ways, e.g., by intravenous, intraperitoneal, subcutaneous, intramuscular, topical or intradermal administration. The dosage regimen will be determined by the attending physician and other clinical factors. As is well known in the medical arts, dosages for any one patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Dosages will vary but a preferred dosage for intravenous administration of DNA is from approximately 10^6 to 10^{22} copies of the nucleic acid molecule. Proteins or peptides may be administered in the range of 0,1ng to 10mg per kg of body weight. The compositions of the invention may be administered locally or systematically. Administration will generally be parenterally, e.g., intravenously; DNA may also be administered directly to the target site, e.g., by biolistic delivery to an internal or external target site or by catheter to a site in an artery.

The present invention further relates to a method for the production of a pharmaceutical composition comprising formulating an inhibitor of said pair or complex of interacting molecules identified by the method of the invention in a pharmaceutically acceptable form.

The inhibitor may be identified according to conventional protocols. Additionally, molecules that inhibit existing protein-protein interactions can be isolated with the yeast two-hybrid system using the URA3 readout system. Yeast cells that express interacting GAL4ad and LexA fusion proteins which activate the URA3 readout system are unable to grow on selective medium containing 5-FOA. However, when an additional molecule is present in these cells which disrupts the interaction of the fusion proteins the URA3 readout system is not activated and the yeast cells can grow on selective medium containing 5-FOA. Using this method potential inhibitors of a protein-protein interaction can be isolated from a library comprising these inhibitors. Systems corresponding to the URA 3 system may be devised by the person skilled in the art on the basis of the teachings of the present invention and are also comprised thereby.

Also, the present invention relates to a method for the production of a pharmaceutical composition comprising identifying a further molecule in a cascade of interacting molecules, of which the at least one member of interacting molecules identified by the method specified herein above is a part of or identifying an inhibitor of said further molecule.

Once at least one member of interacting molecules has been identified, it is reasonable to expect that said member is a part of a biological cascade. Identification of additional members of said cascade can be effected either by applying the method of the present invention or by applying conventional methods. Also, inhibitors of said further members can be identified and can be formulated into pharmaceuticals compositions.

The present invention relates further to a kit comprising:

(i) host cells as identified in any of the preceding claims and at least one genetic element comprising said genetic information specifying at least one of said possibly interacting molecules specified herein above;

(ii) host cells as identified in any of the preceding claims and at least one genetic element not comprising genetic information specifying at least one of said potential interacting molecules specified herein above;

(iii) at least one genetic element comprising said genetic information specifying at least one of said possibly interacting molecules specified herein above;

(iv) at least one genetic element not comprising genetic information specifying at least one of said potential interacting molecules specified herein above;

(v) host cells comprising at least one and preferably at least two of said genetic elements specified in (iii) or (iv);

(vi) at least one planar carrier carrying nucleic acid or protein from said host cells comprising at least one member of said genetic elements specified herein above wherein said nucleic acid or protein is affixed to said carrier in grid form and optionally

solutions to effect hybridization or binding of nucleic acid probes or proteins to said molecules affixed to said grid;

(vii) at least one storage compartment, planar carrier or computer disc comprising or/and characterizing genetic elements, host cells, storage compartments or carriers identified in any of (i) to (vi); and/or

(viii) at least one yeast strain comprising a *ura*3 mutation and a URA3 reporter system.

Preferably, said kit comprises or also comprises at least one storage compartment containing the host cells of (i), (ii) or (v) and/or comprises or also comprises at least one storage compartment containing said genetic information or said potentially interaction molecules encoded by said genetic information as specified in (i) or (iii).

The invention also relates to the use of any of the yeast strains described herein above and in the appended examples for the identification of at least one member of a pair of potentially interacting molecules.

The figures show:

Figure 1

Schematic overview of the 5-FOA counterselection method used to eliminate false positive clones that express single fusion proteins able to auto-activate the *URA3* reporter system. a) Yeast cells expressing a LexA or GAL4ad fusion protein which do not activate the URA3 reporter system can grow in selective medium containing 5-fluoroorotic acid (5-FOA), b) In contrast, yeast strains expressing auto-activating fusion proteins will die when the URA3 reporter system orotidine-5′-phosphate decarboxylase is expressed that converts 5-FOA into the toxic compound 5-fluorouracil which is lethal for yeast cells.

Figure 2

A schematic overview of one embodiment of the method of the invention. For the parallel analysis of a network of protein-protein interactions with the method of the invention, two libraries of plasmid constructs that express DNA binding domain or activation domain fusion proteins are provided. These libraries may consist of specific DNA fragments or a multitude of unknown DNA fragments ligated into binding domain and activating domain plasmids which contain the selectable markers TRP1 and LEU2, respectively. The libraries are transformed into either Mata or Matα yeast strains containing the novel URA3 reporter system and are subsequently plated onto selective media containing 5-fluoroorotic acid (5-FOA). Only those yeast cells that express fusion proteins unable to auto-activate the URA3 reporter system will grow in the presence of 5-FOA. The resulting yeast strains that express only non-auto-activating molecules can then be directly used in an automated interaction mating approach to generate ordered arrays of diploid strains which can be assayed for activation of the lacZ readout system. a) Individual yeast cells that express single fusion proteins unable to activate the URA3 reporter system are transferred into wells of a 384-well microtitre plate using a picking robot. The yeast strains held in the microtitre plates can optionally be replicated and stored. The microtitre plates contain a growth medium lacking amino acids appropriate to maintain the corresponding plasmids in the yeast strains. The interaction matings are subsequently performed by automatically transfering a *Mata* and a *Matα* yeast strain to the same position on a Nylon membrane using automated systems as described by Lehrach *et al.* (1997). Alternatively, a pipetting or micropipetting system (Schober *et al.* 1993) can be used to transfer small volumes of individual liquid cultures of a yeast strain onto a planar carrier onto which a lawn of yeast cells derived from at least one yeast clone of the opposite mating type is sprayed or applied. By both methods ordered arrays of yeast clones are incubated overnight at 30°C to allow interaction mating to occur. The resulting diploid cells are then analysed in a ß-Gal assay as described by Breeden & Nasmyth (1985). b) Yeast strains that grew on selective media containing 5-FOA are pooled and interaction mating between the *Mata* and *Matα* strains is made within liquid

YPD medium. Those diploid yeast strains that express interacting proteins are selected by plating on selective medium lacking histidine and uracil. The selective markers TRP1 and LEU2 maintain the plasmids in yeast strains grown on selective media. HIS3, URA3 and lacZ represent reporter genes in the yeast cells, which are expressed on activation by interacting fusion proteins. The readout system is, in the present case, growth on medium lacking histidine and/or uracil and enzymatic activity of ß-galactosidase which can be screened at a later time point. A colony picking robot is used to pick the diploid yeast colonies into individual wells of 384-well microtiter plates containing selective medium lacking histidine, and the resulting plates are incubated at 30°C to allow cell growth. The interaction library optionally may be replicated and stored. Using a spotting robot, diploid cells are transferred to replica membranes which are subsequently placed onto growth medium. The resulting regular arays of diploid yeast clones are analysed for ß-Gal activity as described by Breeden & Nasmyth (1985). In either case a) and b), a digital image from each dried membrane is obtained with a CCD camera which is then stored on computer. Using digital image procesing and analysis (Lehrach et al. 1997) clones that express interacting fusion proteins can be identified by considering the ß-Gal activity of these clones spotted in a defined pattern on the membrane. The individual members comprising the interactions can then be identified by one or more techniques, including PCR, sequencing, hybridisation, oligofingerprinting or antibody reactions.

Figure 3

The applications of an established and exhaustive network of protein-protein interactions. The identity of positive clones and the identity of the members comprising the interactions for the entire interaction libray can be stored in a database. These data are used to establish a network of protein-protein interactions which can be used for a variety of purposes. For example, to predict the existance of new biological interactions or pathways, or to determine links between biological networks. Furthermore with this method, the function and localisation of previously unknown proteins can be predicted by determining their interaction partners. It also can be used to predict the

response of a cell to changes in the expression of particular members of the networks. Finally, these data can be used to identify proteins within a medically relevant pathway which are suitable for theraputic,diagnosis intervention and for the treatment of disease.

Figure 4

The structure of the URA3 reporter system carried by the plasmid pLUA. Important features of pLUA include the URA3 gene which is under the transcriptional control of the lexAop-GAL1 promoter, the *ADE2* selectable marker that allows yeast *ade2*-auxotrophs to grow on selective media lacking adenine and the ß-lactamase gene (bla) which confers ampicilin resistance in E.coli. The pLUA plasmid replicates automously both in yeast using the 2µ replication origin and in E.coli using the ColE1 origin of replication.

Figure 5

Results of automated interaction mating to identify diploid yeast strains that express interacting fusion proteins. a) Progeny of the yeast strains x1a and x2a were spotted at positions 1 and 2 on a nylon membrane using a spotting robot such as described by Lehrach et al. (1997). The yeast strains y1α and y2α of the opposite mating type were subsequently spotted on positions 1 and 2 which already contained cells from the strains x1a and x2a. To assist in recognition of the duplicate spotting pattern, ink was spotted in position 2 directly to the right of the spotted yeast clones. b) The membrane was transferred to a YPD agar plate and was incubated at 30° C overnight to allow interaction mating to occur. c) Diploid yeast cells that had grown on the membrane were subsequently analysed for ß-galactosidase activity using the method of Breeden & Nasmyth (1985).

The example illustrates the invention:

Example 1. Construction of a novel URA3 reporter system to eliminate false positive yeast clones, and its use in an automated interaction mating approach of the yeast 2 hybrid system

Figure 4 shows the novel URA3 reporter system carried by the plasmid pLUA. This unique URA3 reporter system under the control of a bacterial LexAop upstream activation sequence (UAS) can be used within the yeast 2-hybrid system both as a counterselectable marker and as a positive selection marker to eliminate false positive clones. The plasmid contains the unique features of the UAS_{lexAop}-URA3 reporter system, the selecable marker ADE2 that allows yeast ade2-auxotrophs to grow on selective media without adenine and the *bla* gene which confers amplicillin resistance in *E.coli*. The plasmid pLUA is a shuttle vector that replicates autonomously in *E. coli* and yeast.

For the construction of pLUA a 1.5 kb *Sac* I/*Cla* I UAS_{lexAop}-URA3 fragment was isolated from pBS-lexURA and ligated together with a 2.4 kb *Sac* I/*Cla* I *ADE2* fragment into *Cla* I digested pGAD425Δ. pBS-lexURA was generated by ligating a 0.88 kb *Hind* III/*Bam* HI URA3 fragment isolated from pCR2.1-. URA3 together with a *Sac* I/*Bam* HI UAS_{lexAop} fragment isolated from pCR2.1-UAS_{lexAop} into *Sac* I/*Bam* HI digested pBluescript SK+. The URA3 and UAS_{lexAop} fragments were obtained by PCR using genomic DNA from *S. cerevisiae* strain L40c and were cloned into a pCR2.1 vector (Invitrogen). The ADE2 gene was isolated by PCR using genomic DNA from SEY6210α. pGAD425Δ was generated by deleting of an 1.2 kb *Sph* I fragment from pGAD425 (Han and Colicelli, 1995) and religation of the vector.

To test the functionality of the URA3 reporter construct present in pLUA, two yeast strains, both containing a *ura3* mutation were generated; L40cu (Genotype: *Mata his3Δ200 trp1-910 leu2-3,112 ade2 LYS2::(lexAop)$_4$-HIS3 ura3::(lexAop)$_8$-lacZ Gal4 can1*); and AMR70u (Genotype: *Matα his3Δ200 lys2-801am trp1-910 leu2-3,112 ade2 ura3::(lexAop)$_8$-lacZ*). L40cu and AMT70u are isogenic to L40c and AMR70, respectively, except that each

strain carries a *ura3* mutation (Wanker et al., 1997; Hollenberg et al., 1995). These mutants were selected by plating L40c and AMR70 cells on minimal medium containing 0.2 % 5-fluoroorotic acid (5-FOA) (Sigma).

Yeast cells that express the URA3 gene are sensitive to the uracil analogue 5-fluoroorotic acid (5-FOA) and cannot grow on selective media containing 5-FOA since the enzyme encoded by the URA3 gene metabolises 5-FOA to 5-fluorouracil, which is toxic to yeast cells (Boeke *et al.*, 1984). It is well known that some LexA or GAL4ad fusion proteins when expressed in the yeast strain L40c, are able to auto-activate the reporter genes LacZ and HIS3 without the need for any interacting fusion proteins. These false positive clones can now be eliminated with the novel URA3 reporter system by plating cells on selective medium containing 5-FOA (Figure 1).

It is well known that yeast interaction mating can be used with the yeast 2 hybrid system to combine within a single diploid strain, plasmids that encode for both LexA and GAL4ad fusion proteins (Hollenberg et al., 1995). By mating strains that are able to grow on 5-FOA selective media, only fusion proteins unable to auto-activate the novel URA3 reporter system will be tested for their ability to interact with other fusion proteins. Interaction mating between a AMR70u yeast strain and a L40cu strain which each contains a genetic element expressing a LexA fusion protein or a genetic element expressing a GAL4ad fusion protein that is unable to auto-activate the URA3 reporter system, it should be possible to produce a diploid yeast strain that expresses both the LexA fusion and Gal4ad fusion proteins. If these two fusion proteins interact, the readout system should be activated, and can be assayed by testing for expression of the HIS3, LacZ or URA3 genes. Thus, yeast clones that express interacting GAL4 and LexA fusion proteins can be detected by their ability to express the LacZ gene and turn blue on incubation with X-Gal.

To examine whether the novel URA3 reporter system can be used to eliminate false positive clones within an automated interaction mating approach of the yeast two-hybrid system as shown in Figure 2, a series of

plasmids that express various LexA and GAL4ad fusion proteins were constructed as listed in Table 1. It was shown that the proteins LexA, LexA-SIM1, GAL4ad and GAL4ad-ARNT when expressed alone, were unable to activate a LacZ and the HIS3 readout system. (Probst *et al.*, 1997; Vojtek *et al.*, 1993). However, when the fusion proteins LexA-SIM1 and GAL4ad-ARNT are expressed together in one cell they are capable of activating the readout system (Probst *et al.*, 1997). In comparison, the fusion proteins LexA-HIP1 and GAL4ad-LexA encoded by the plasmids pBTM117c-HIP1 and pGAD425-LexA, respectively, without the need for any interacting partner proteins were capable of activating the same reporter genes. Thus, these fusion proteins should activate the novel URA3 reporter gene in a similar manner since it also contains the same LexAop upstream activation sequence as the HIS3 and LacZ readout system.

Three mating type-a yeast strains were constructed by cotransformation into L40cu, of the plasmid pLUA containing the URA3 reporter system, and either the pBTM117c, pBTM117c-SIM1 or pBTM117c-HIP1 plasmids respectively. Transformants that contained both the pLUA plasmid and one of the DNA binding domain plasmid were selected on SD-trp-ade medium. Three mating type-α yeast strains were similarly constructed by cotransformation into AMR70u of pLUA, and either the pGAD425, pGAD425-ARNT or pGAD425-lexA plasmids respectively. Transformants that contained both the pLUA and one of the activation domain plasmids were selected on SD-leu-ade medium. The yeast strains thus obtained are listed in Table 2. The yeast strains x1a, x2a and x3a were replica plated onto the selective media SD-trp-ade, SD-trp-ade containing 0.2% 5-FOA and SD-trp-ade-ura, while the yeast strains y1α, y2α and y3α were replica plated onto the selective media SD-leu-ade, SD-leu-ade containing 0.2% 5-FOA and SD-leu-ade-ura. Table 3 shows that the two yeast strains x3a and y3α which expressed the fusion proteins LexA-HIP1 and GAL4ad-LexA respectively were unable to grow on their respective media containing 5-FOA yet were able to grow on their respective media lacking uracil. In contrast, all other yeast strains which contained plasmids that expressed fusion proteins which were alone unable to activate the reporter system could grow on their respective media containing 5-FOA, but could not

grow on selective media lacking uracil. This indicates that it is possible to eliminate yeast clones that express single fusion proteins which auto-activate the reporter system, by selection on media containing 5-FOA.

The yeast strains that did not express auto-activating fusion proteins were mated using an automated approach as follows. Each of the yeast strains x1a, x2a, y1α and y2α was grown in every well of one of four microtitre plates containing SD-trp-ade medium for the *Mat*a strains and SD-leu-ade medium for the *Mat*α strains. Using a spotting robot such as described by Lehrach *et al.* (1997), the yeast strains x1a and x2a were transfered in a defined 2 x 2 duplicate pattern to Nylon membrane (Amersham) which had been pre-soaked with YPD medium. Using the same automated system, the yeast strains y1α and y2α were then transfered to the same respective spotting positions on each membrane as, and already containing the x1a and x2a clones. The spotted membranes were transfered to YPD plates and incubated for over night at 30°C to allow mating and growth to occur. Each membrane was assayed for ß-Gal activity using the method of Breeden & Nasmyth (1985) and was subsequently air dried overnight. A digital image of each dried filter was obtained using a high resolution charge coupled devive (CCD) camera, the image was transferred to computer and the identity of clones that expressed ß-Gal was determined using a semi-automated image analysis system such as described by Lehrach *et al.* (1997).

Figure 5 shows the results of automated interaction mating between the strains x1a & y1α and x2a & y2α. Both resulting diploid strains grew on YPD media, yet only the diploid strain resulting from the interaction mating of x2a & y2α that contained plasmids encoding the interacting fusion proteins LexA-SIM1 & GAL4ad-ARNT respectively, showed a LacZ+ phenotype and turned blue on incubation with X-Gal. No ß-galactosidase activity was observed for the diplod strain resulting from the interaction mating between the strains x1a and y1α that contained plasmids encoding the proteins LexA and GAL4ad.

Thus, these results indicate that the novel *URA3* reporter system can be used to eliminate clones containing auto-activting fusion proteins prior to interaction

mating. Secondly, using an automated interaction mating approach it is possible to effeciently test for interaction using the yeast 2-hybrid system between only those fusion proteins unable to activate the novel URA3 reporter system.

Table 1: Two-hybrid vectors used for the expression of fusion proteins.

plasmid	fusionprotein	insert (kb)	selection in yeast	reference
pBTM117c	LexA	-	TRP1	Wanker et al. 1997
pBTM117c-SIM1	LexA-SIM1	1.1	TRP1	Probst et al., 1997
pBTM117c-HIP1	LexA-HIP1	1.2	TRP1	Wanker et al., 1997
pGAD425	GAL4ad	-	LEU2	Han & Colicelli, 1995
pGAD425-ARNT	GAL4ad-ARNT	1.3	LEU2	Probst et al., 1997
pGAD425-lexA	GAL4ad-LexA	0.6	LEU2	this work

Table 2: Yeast strains used for the 5-FOA counterselection and the automated interaction mating

strain	plasmids	selected on
x1a	pBTM117c / pLUA	SD-trp-ade
x2a	pBTM117c-SIM1 / pLUA	SD-trp-ade
x3a	pBTM117c-HIP1 / pLUA	SD-trp-ade
y1α	pGAD425 / pLUA	SD-leu-ade
y2α	pGAD425-ARNT / pLUA	SD-leu-ade
y3α	pGAD425lexA / pLUA	SD-leu-ade

Table 3. Identification of fusion proteins that activate the URA3 readout system.

a.

strain	SD-trp-ade	SD-trp-ade+5-FOA	SD-trp-ade-ura
x1a	+	+	-
x2a	+	+	-
x3a	+	-	+

SD-trp-ade: Selective medium lacking tryptophan and adenine.
SD-trp-ade+5-FOA: Selective medium containing 0.2 % 5-FOA.
SD-trp-ade-ura: Selective medium lacking tryptophan, adenine and uracil.

b.

strain	SD-leu-ade	SD-leu-ade+5-FOA	SD-leu-ade-ura
y1α	+	+	-
y2α	+	+	-
y3α	+	-	+

SD-leu-ade: Selective medium lacking leucin and adenine.
SD-leu-ade+5-FOA: Selective medium containing 0.2 % 5-FOA.
SD-leu-ade-ura: Selective medium lacking leucin, adenine and uracil.

References

Bartel, P., Chien, C.-T., Sternglanz, R. and Fields, S. (1993). Elimination of false positives that arise in using in the two-hybrid system. *BioTechniques* **14**: 920-924

Bartel, P.L., Roecklein, J.A., SenGupta, D., and Fields, S. (1996). A protein linkage map of
Escherichia coli bacteriophage T7. *Nat. Genet.* **12**: 72-77.

Boeke, J.D., LaCroute, F. and Fink, G.R. (1984). A positive selection for mutants lacing orotidine-5'-phosphate decarboxylase activitiy in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* **197**: 345-346

Breeden, L. and K. Nasmyth, K. (1985). Regulation of yeast HO gene. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 643-650.

Cubbitt, A.B., Heim, R., Adams, S.R., Boyd, A.E., Gross, L.A. and Tsien, R.Y. (1995). Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **20**: 448-455.

DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent J.M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**: 457-460.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction nuclease fragments to high specific activity. *Anal. Biochem.* **132**: 6-13

Fields, S.and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245-246.

Han, L. and Colicelli, J. (1995). A human protein selected for interference with Ras function interacts directly with Ras and competes with Raf1. *Mol. Cell. Biol.* **15**: 1318-1323.

Hollenberg, S.M., Sternglanz, R., Cheng, P.F. and Wintraub, H. (1995). Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Mol. Cell. Biol.* **15**: 3813-3822

Larin, Z and Lehrach, H., (1990). Yeast artificial chromosomes : an alternative approach to the molecular analysis of mouse developmental mutations. *Genet. Res.* **56:** 203-208

Lehrach, H., Bancroft, D. and Maier, E. (1997). Robotics, computing, and biology: An interdisciplinary approach to the analysis of complex genomes. *Interdisp. Science Rev.* **22**: 37-43.

Li, J.J. and Herskowitz, I (1993). Isolation of ORC6, a component of the yeast origin recognition complex by a one hybrid system. *Science* **262**: 1870-1874.

Lockhart, D. J., Dong, H. Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E. L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* 12: 1675-1680.

Maier, E., Meier-Ewert, S., Bancroft, D. and Lehrach H. (1997). Automated array technology for gene expression profile. *Drug Disc. Today* **2**: 315-324.

Pansegrau, W., Miele, L., Lurz, R. and Lanka E. (1987). Nucleotide sequence of the kanamycin resistance determinant of plasmid RP4: homology to other aminoglycoside 3'-phosphotransferases. *Plasmid* **18**: 193-204.

Probst, M.R., Fan, C.M., Tessier-Lavigne, M. and Hankinson, O. (1997). Two murine homologs of the Drosophila single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein. *J. Biol. Chem.* **272**: 4451-4457.

Putz, U., Skehel, P. and Kuhl, D. (1996). A tri-hybrid system for the analysis and detection of RNA--protein interactions. *Nucleic Acids Res.* **24**: 4838-4840

Schober, A., Guenther, R. Schwienhorst, A., Doering, M. and Lindemann, B. F. (1993). Accurate high-speed liquid handling of very small biological samples. Biotechniques 15: 324-329.

SenGupta DJ. Zhang B. Kraemer B. Pochart P. Fields S. Wickens M. (1996). A three-hybrid system to detect RNA-protein interactions in vivo. *PNAS*. **93**: 8496-501

Shalon, D., Smith, S. J. and Brown, P. O. (1996). A DNA microarray system for analysing complex DNA samples using two-colour fluorescent probe hybridization. *Genome Res.* 6, 639-645.

Vojtek, A.B., Hollenberg, S.M. and Cooper, J.A. (1993).Mammalian Ras interacts directly with serine/threonine kinase Raf. *Cell* 74: 205-214.

Wanker, E.E., Rovira, C., Scherzinger, E., Hasenbank, R., Waelter, S., Tait, D., Colicelli, J. and Lehrach H. (1997). HIP-I: a huntingtin interacting protein isolated by the yeast two-hybrid system. *Hum. Mol. Genet.* **6**: 487-495.

Wu, L.C., Wang, Z.W., Tsan, J.T., Spillman, M.A., Phung, A., Xu, X.L., Yang, M.C., Hwang, L.Y., Bowcock, A.M. and Baer, R. (1996). Identification of a RING protein that can interact in vivo with the BRCA1 gene product. *Nat. Genet.* **14**: 430-440.

Yang. M., Wu, Z. and Fields (1995). Protein-peptide interaction analyzed with the yeast two-hybrid system. *Nucleic Acids Res.* 1152-1156

Claims

1. A method for the identification of at least one member of a pair or a complex of interacting molecules, comprising:

(a) providing at least two sets of host cells each containing at least one genetic element with a selectable marker different for each set of host cells, said genetic elements each comprising genetic information specifying one of said molecules, said host cells further carrying a readout system that is activated upon the presence of autoactivating molecules;

(b) selecting against molecules that auto-activate said readout system by transferring progeny of at least one set of host cells to:

(ba) a selective medium which allows growth of said host cells in the presence of said selectable marker different for each set of host cells and which precludes growth of said host cells upon auto-activation of said readout system; or

(bb) a selective medium which allows growth of all of said host cells and visual differentiation between those cells whose readout system has been auto-activated and those whose readout system has not been auto-activated;

(c) combining in host cells said genetic elements from at least two different sets of host cells, wherein at least one set of host cells grows on said selective medium specified in (ba) or does not auto-activate said readout system on said selective medium specified in (bb);

(d) allowing at least one interaction, if any, to occur;

(e) identifying host cells obtained in step (c) or (d) containing interacting molecules that activate said readout system:

(f) identifying at least one member of said pair or complex of interacting molecules.

2. The method of claim 1 wherein said pair or complex of interacting molecules is selected from the group consisting of RNA-RNA, RNA-

DNA, RNA-protein, DNA-DNA, DNA-protein, protein-protein, protein-peptide, or peptide-peptide interactions.

3. The method of claim 1 or 2 wherein said genetic elements are plasmids, artificial chromosomes, viruses or other extra chromosomal elements.

4. The method of any one of claims 1 to 3 wherein said interactions lead to the formation of a functional transcriptional activator comprising a DNA-binding and a transactivating protein domain and which is capable of activating a responsive moiety that drives the activation of said readout system.

5. The method of claim 4 wherein said readout system is a detectable protein.

6. The method of claim 5 wherein said detectable protein is encoded from at least one of the genes lacZ, HIS3, URA3, CAN1, LYS2, CYH2, sacB or HPRT.

7. The method of any one of claims 1 to 6 wherein said host cells are yeast cells, bacterial cells, mammalian cells, insect cells or plant cells.

8. The method of any one of claims 1 to 7 further comprising transforming or transfecting said sets of host cells with said genetic elements prior to step (a).

9. The method of any one of claims 1 to 8 wherein cell fusion, conjugation or interaction mating is used for the generation of said host cells with said genetic elements in step (c).

10. The method of any one of claims 1 to 9 wherein said visual differentiation in step (bb) is based on a color reaction which is induced by said readout system.

11. The method of any one of claims 1 to 10 wherein said selectable marker is an auxotrophic or antibiotic marker.

12. The method of claim 11 wherein said auxotrophic or antibiotic marker is LEU2, TRP1, URA3, ADE2, HIS3, LYS2 or Zeocin.

13. The method of any one of claims 1 to 12 wherein said selective medium specified in step (ba) comprises 5-fluoroorotic acid, canavanine, cycloheximide or α-amino adipate.

14. The method of any one of claims 1 to 13 wherein progeny of said host cells of step (b) and/or (d) are transferred to a storage compartment.

15. The method of claim 14 wherein said transfer is effected or assisted by automation or a picking robot.

16. The method of claim 14 or 15 wherein said storage compartment comprises an anti-freeze agent.

17. The method of claim 16 wherein said anti-freeze agent is glycerol which is preferably present in said media in an amount of 3 - 25% (vol/vol).

18. The method of any one of claims 14 to 17 wherein said storage compartment is a microtiter plate.

19. The method of claim 18 wherein said microtiter plate comprises 384 wells.

20. The method of any one of claims 1 to 19 wherein prior to step (c) or (e) said host cells are transferred to a planar carrier which at a later time point is applied to identification step (e).

21. The method of claim 20 wherein the transfer is made or assisted by automation, a spotting robot or high-throughput pipetting device.

22. The method of claims 20 or 21 wherein the transfer is in a regular grid pattern of densities of 1 to 1000 clones per cm^2.

23. The method of any one of claims 20 to 22 wherein said planar carrier is a membrane.

24. The method of any one of claims 1 to 23 wherein said identification of said host cells in step (e) is effected by visual means from consideration of the activation state of said readout system.

25. The method of any one of claims 1 to 24 wherein said identification of said host cells in step (e) is effected by digital image storage, analysis or processing.

26. The method of any one of claims 1 to 25 wherein said identification of said members of said pair or complex of interacting molecules are effected by nucleic acid hybridization, antibody binding or nucleic acid sequencing.

27. The method of claim 26 wherein said identification made by said antibody reaction or said hybridization is effected using regular grids of said at least one member or of said genetic information encoding at least one of said members.

28. The method of claim 27 wherein construction of said regular grids is effected by automation or a spotting robot.

29. The method of any one of claims 26 to 28 wherein said identification is effected by digital image storage, processing or analysis.

30. The method of claim 26 wherein nucleic acid molecules, prior to said identification, are amplified by PCR or are amplified in as a part of said genetic elements, preferably in bacteria and most preferably in E. coli.

31. A method for the production of a pharmaceutical composition comprising formulating said at least one member of said pair or complex of interacting molecules identified by the method of any one of claims 1 to 30 in a pharmaceutically acceptable form.

32. A method for the production of a pharmaceutical composition comprising formulating an inhibitor of the interaction of said pair or complex of interacting molecules identified by the method of any one of claims 1 to 30 in a pharmaceutically acceptable form.

33. A method for the production of a pharmaceutical composition comprising identifying a further molecule of a cascade of interacting molecules, of which the at least one member of said pair or complex of interacting molecules identified by the method of any one of claims 1 to 30 is a part of or identifying an inhibitor of said further molecule.

34. Kit comprising:

(i) host cells as identified in any of the preceding claims and at least one genetic element comprising said genetic information specifying at least one of said possibly interacting molecules specified in any of the preceding claims;

(ii) host cells as identified in any of the preceding claims and at least one genetic element not comprising genetic information specifying at least one of said potential interacting molecules specified in any of the preceding claims;

(iii) at least one genetic element comprising said genetic information specifying at least one of said possibly interacting molecules specified in any of the preceding claims;

(iv) at least one genetic element not comprising genetic information specifying at least one of said potential interacting molecules specified in any of the preceding claims;

(v) host cells comprising at least one and preferably at least two of said genetic elements specified in (iii) or (iv);

(vi) at least one planar carrier carrying nucleic acid or protein from said host cells comprising at least one member of said genetic elements specified in any of the preceding claims wherein said nucleic acid or protein is affixed to said carrier in grid form and optionally solutions to effect hybridization or binding of nucleic acid probes or proteins to said molecules affixed to said grid;

(vii) at least one storage compartment, planar carrier or computer disc comprising or/and characterizing genetic elements, host cells, storage compartments or carriers identified in any of (i) to (vi); and/or

(viii) at least one yeast strain comprising a *ura*3 mutation and a URA3 reporter system.

35. The kit of claim 34 wherein said host cells of (i), (ii) or (v) are contained in at least one storage compartment.

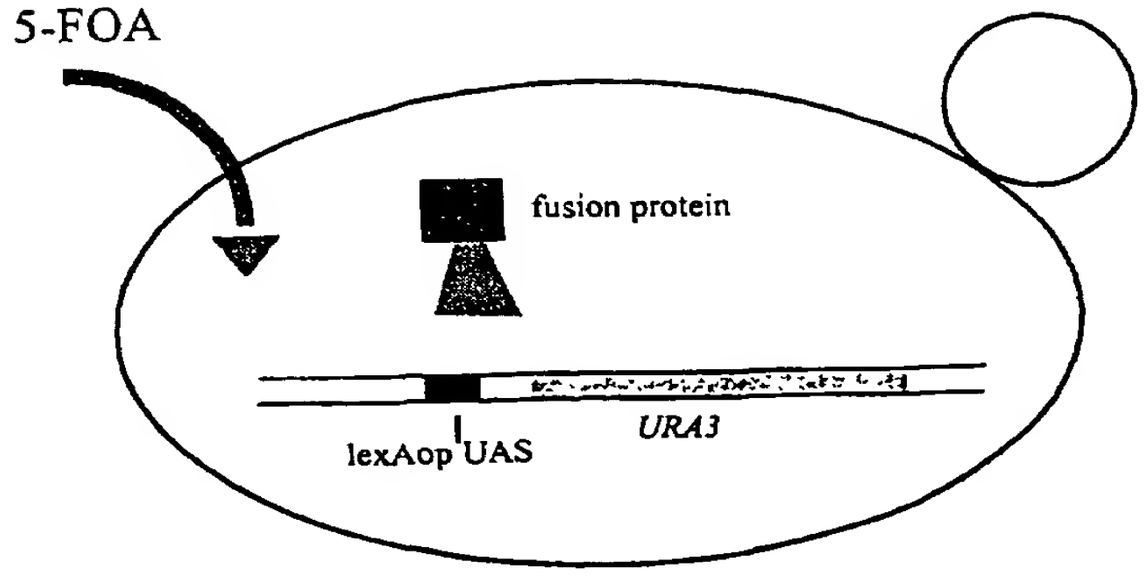
36. The kit of claim 34 or 35 wherein said genetic information or said potentially interaction molecules encoded by said genetic information as specified in (i) or (iii) is contained in at least one storage compartment.

ABSTRACT

The present invention relates to an improved method for the identification and optionally the characterization of interacting molecules designed to eliminate the rather large numbers of false positive clones isolated by conventional two-hybrid systems. The method of the invention relies on a novel combination of selection steps used to detect clones that express interacting molecules. The present invention further relates to a kit useful for carrying out the method of the invention. The present invention provides for parallel, high-throughput or automated interaction screens for the reliable identification of interacting molecules by interaction mating.

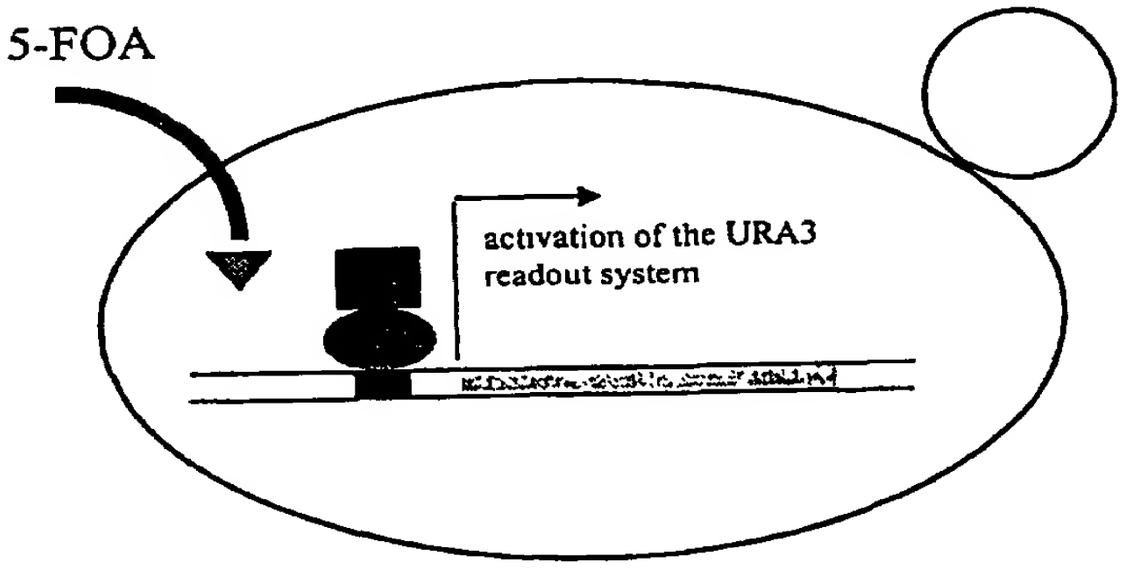
growth

a



+

b



−

5-fluoroorotic acid —orotidine-5′-phosphate decarboxylase (URA3)→ 5-fluorouracil (toxic compound

Figure 1

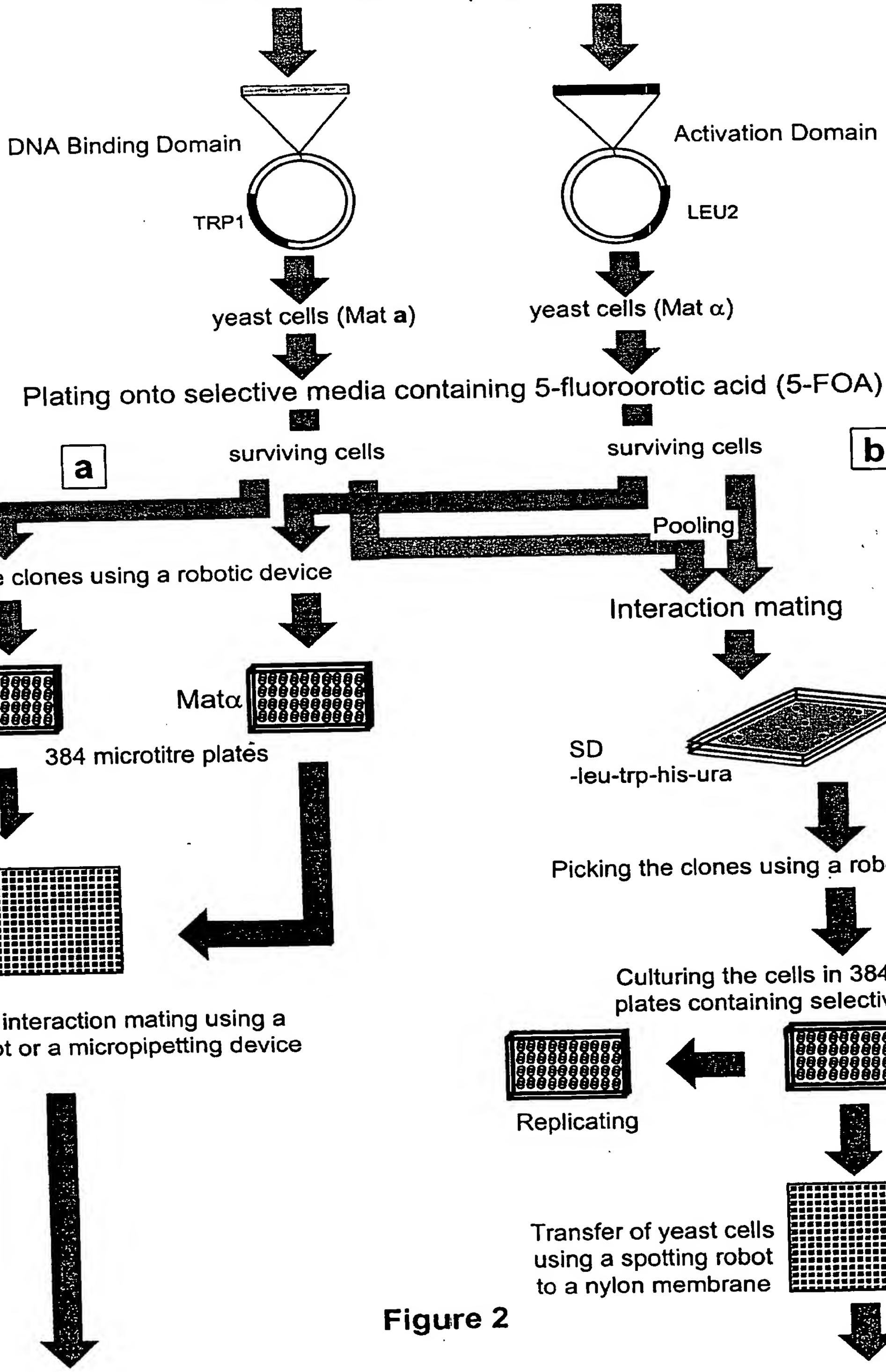
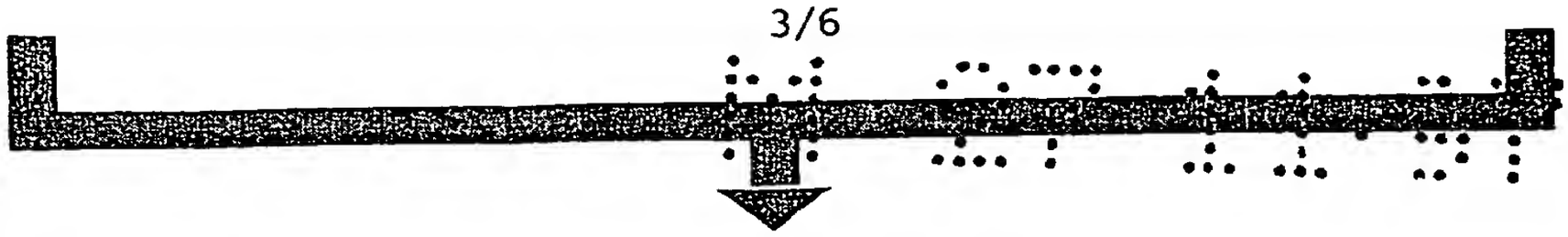
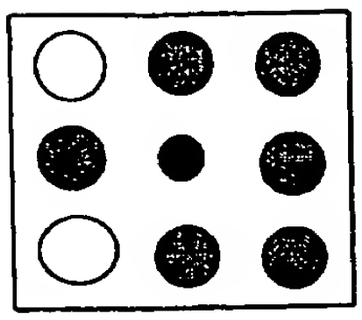


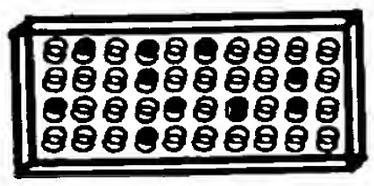
Figure 2

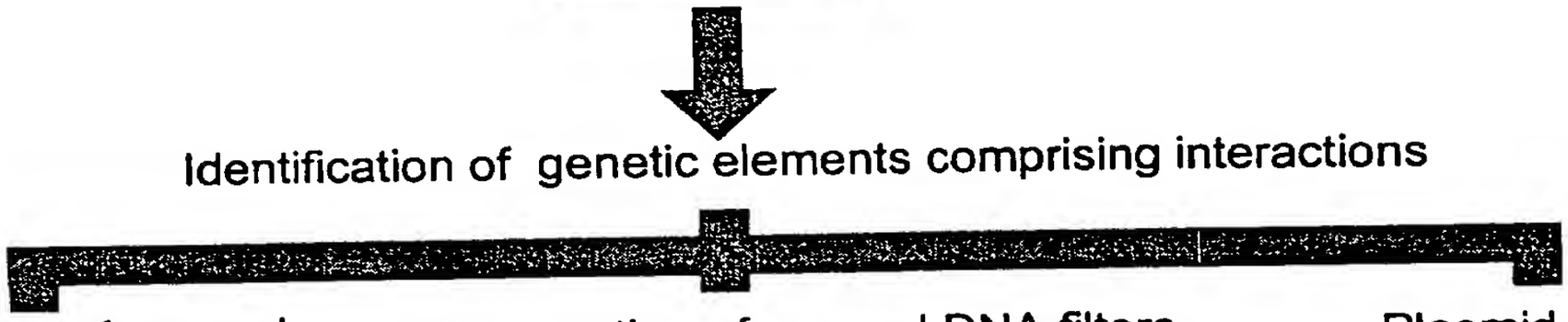


ß-Gal-assay



Identification of clones containing interactions

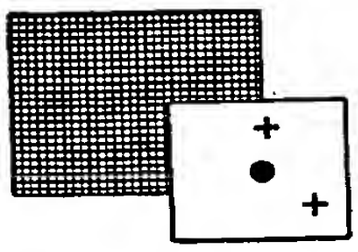




Identification of genetic elements comprising interactions

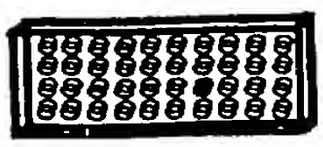
Preparation of arrayed protein filters

Antibody reaction



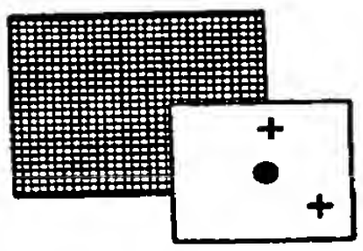
Digital Image processing & Analysis

Identification of interacting molecules and localisation to clone

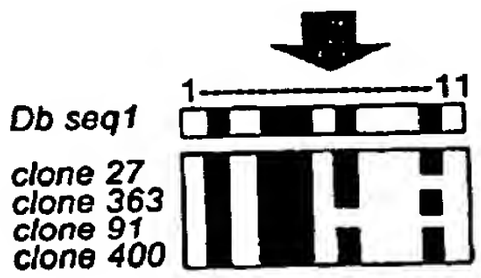


Preparation of arrayed DNA filters

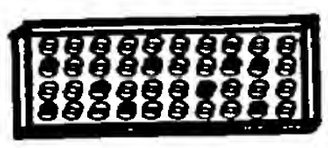
Hybridization using oligonucleotide probes



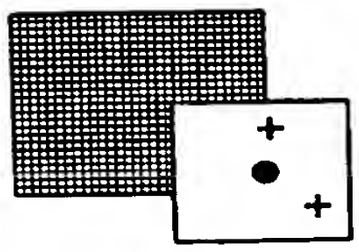
Digital Image processing & Analysis



Identification of clones with a specific fingerprinting pattern

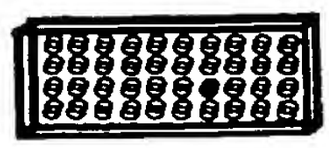


Hybridization using specific DNA-probes



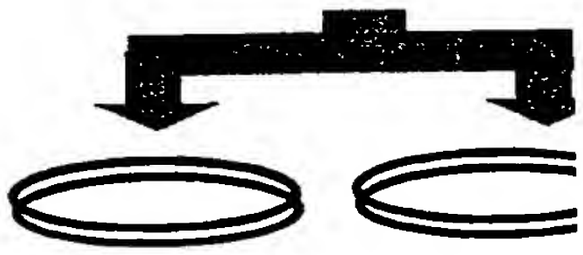
Digital Image processing & Analysis

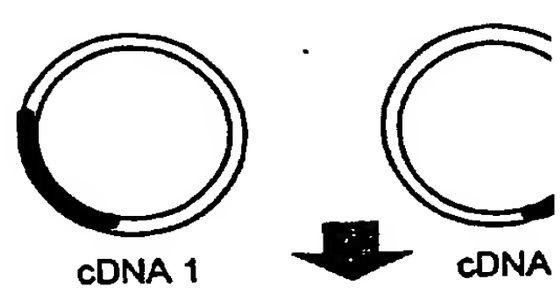
Identification of interacting molecules and localisation to clone



Plasmid isolation

Transformation into I





Sequencing

Figure 2 (continued)

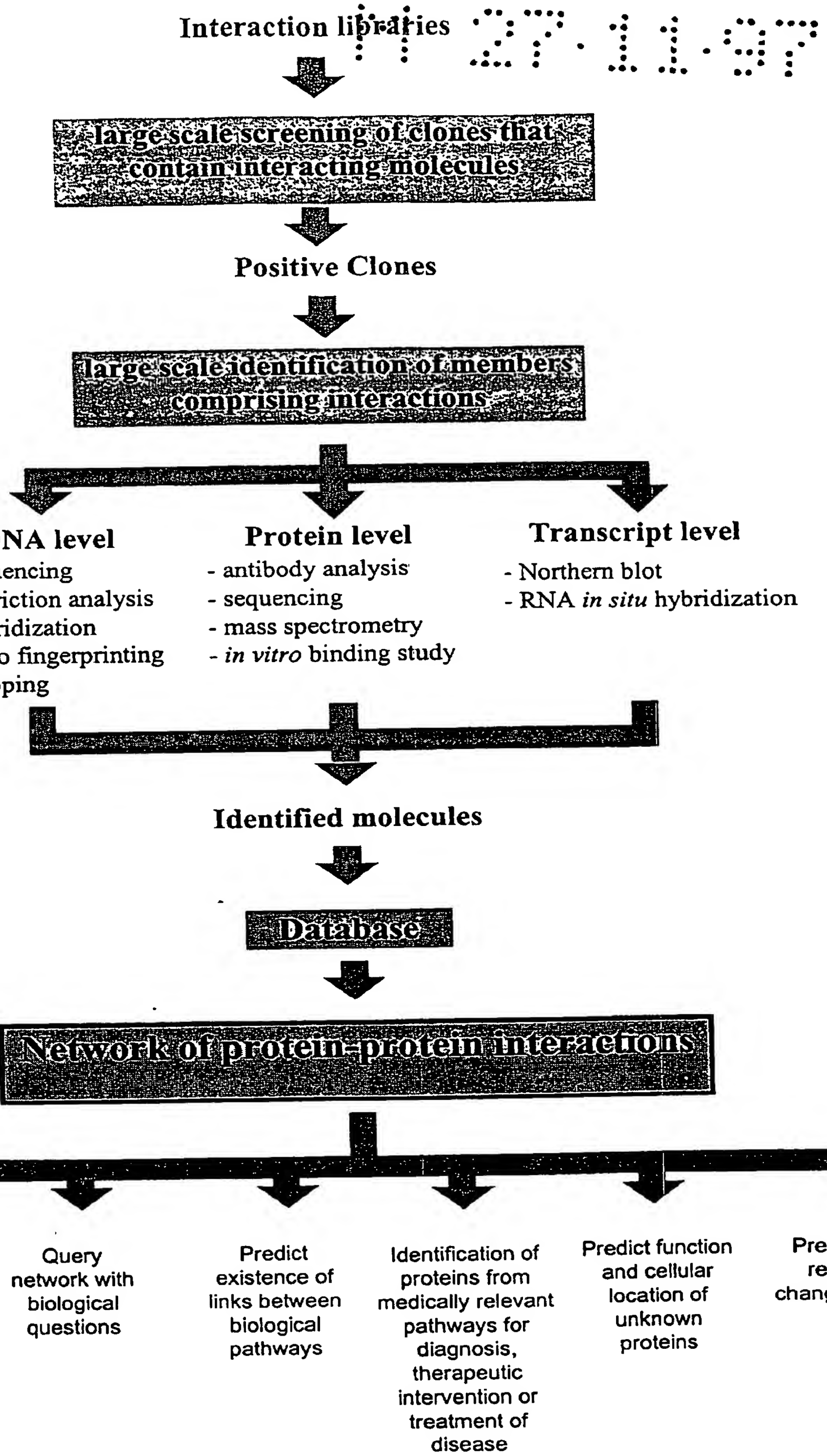


Figure 3

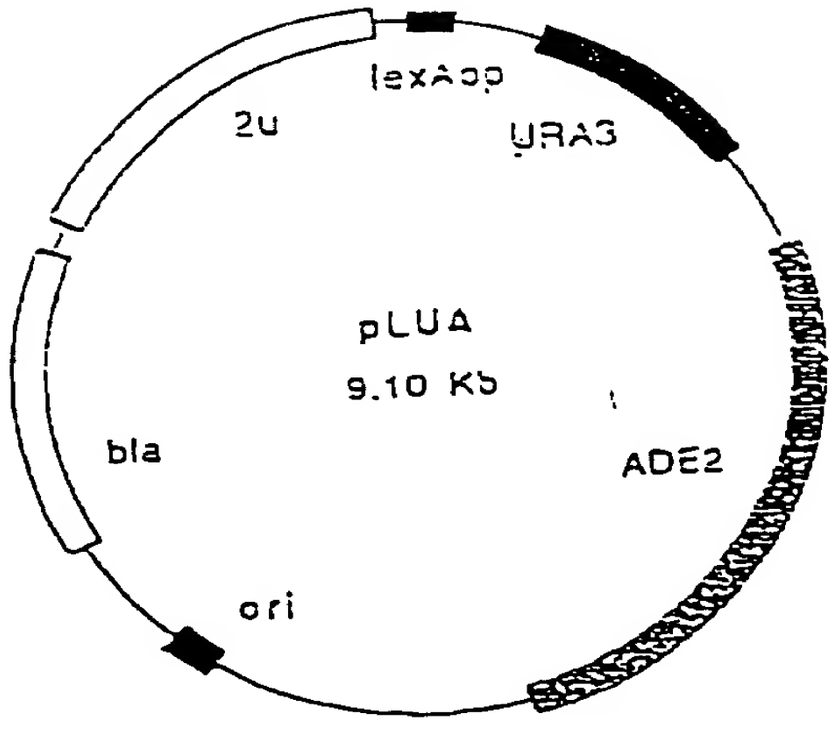


Figure 4

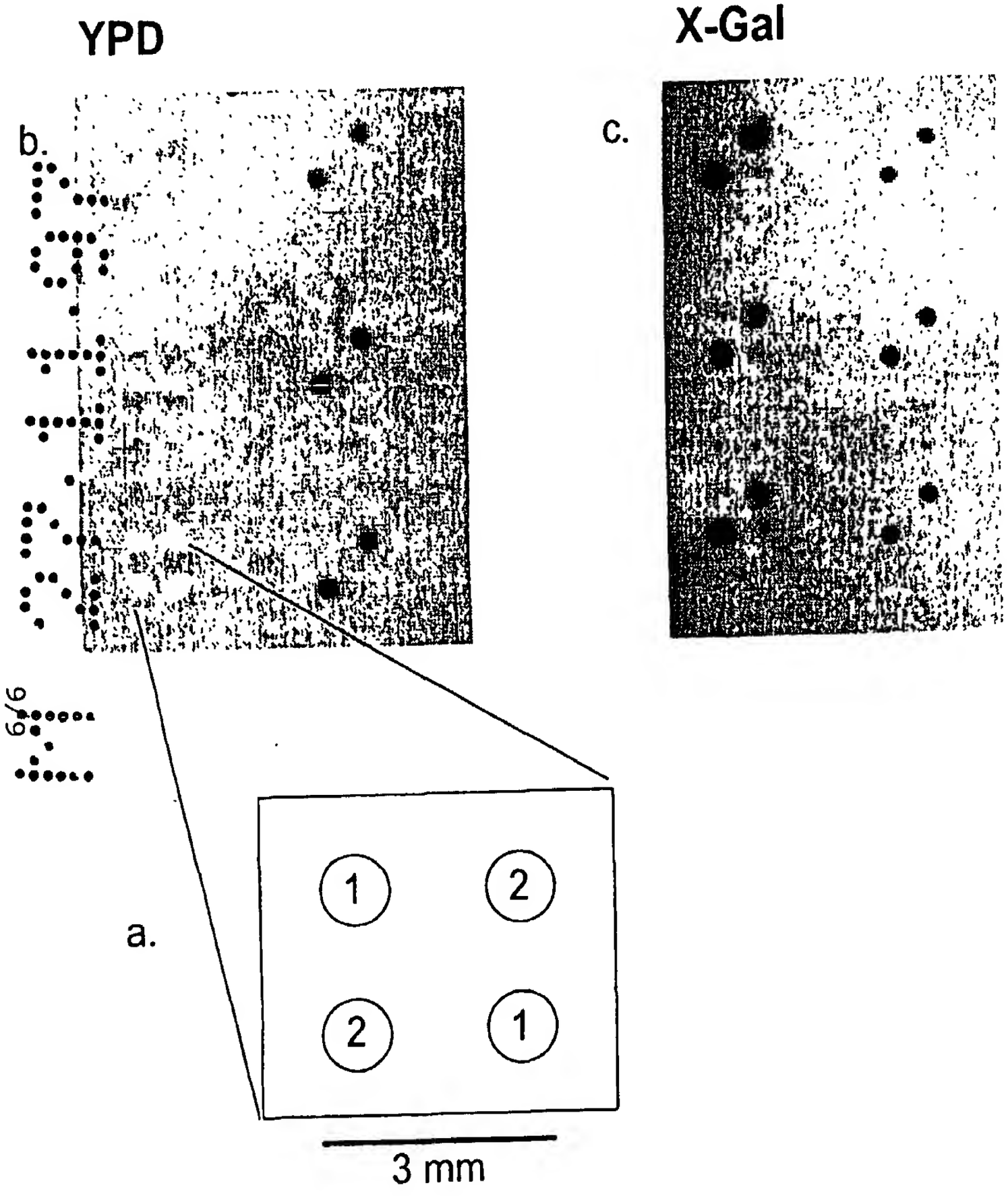


Figure 5